A Methodology for Extracting Standing Human Bodies from Single Images

Dr. Y. Raghavender Rao[1], N. Devadas Naik[2]
[1]*Head ECE JNTUHCEJ Jagtityal*
[2]*Asst professor Sri Chaitanya engineering college*

Abstract: Extraction of the image of human body in unconstrained still images is challenging due to several factors, including shading, image noise, occlusions, background clutter, the high degree of human body deformability, and the unrestricted positions due to in and out of the image plane rotations. we propose a bottom-up approach for human body segmentation in static images. We decompose the problem into three sequential problems: Face detection, upper body extraction, and lower body extraction, since there is a direct pair wise correlation among them.

Index Terms: *Skin segmentation, Torso, Face recognition, Thresholding, Ethnicity, Morphology.*

INTRODUCTION

In this study, we propose a bottom-up approach for human body segmentation in static images. We decompose the problem into three sequential problems: Face detection, upper body extraction, and lower body extraction, since there is a direct pair wise correlation among them. Face detection provides a strong indication about the presence of humans in an image, greatly reduces the search space for the upper body, and provides information about skin color. Face dimensions also aid in determining the dimensions of the rest of the body, according to anthropometric constraints. This information guides the search for the upper body, which in turns leads the search for the lower body. Moreover, upper body extraction provides additional information about the position of the hands, the detection of which is very important for several applications. The basic units upon which calculations are performed are super pixels from multiple levels of image segmentation. The benefit of this approach is twofold. First, different perceptual groupings reveal more meaningful relations among pixels and a higher, however, abstract semantic representation. Second, a noise at the pixel level is suppressed and the region statistics allow for more efficient and robust computations. Instead of relying on pose estimation as an initial step or making strict pose assumptions, we enforce soft anthropometric constraints to both search a generic pose space and guide the body segmentation process. An important principle is that body regions should be comprised by segments that appear strongly inside the hypothesized body regions and weakly in the corresponding background. Without making any assumptions about the foreground and background, except for the assumptions that sleeves are of similar color to the torso region, and the lower part of the pants is similar to the upper part of the pants, we structure our searching and extraction algorithm based on the premise that colors in body regions.

We classify approaches for human body segmentation into the following categories. The first includes interactive methods that expect user input in order to discriminate the foreground and background. Interactive segmentation methods are useful for generic applications, and have the potential to produce very accurate results in complex cases. The second category includes top-down approaches, which are based upon a priori knowledge, and use the image content to further refine an initial model. Top-down approaches have been proposed as solutions to the problem of segmenting human bodies from static images. The main characteristic of these approaches is that they require high-level knowledge about the foreground, which in the case of humans is their pose. One method for object recognition and pose estimation is the pictorial structures (PS) model and its variations. In general, human body segmentation approaches based on PS models can deal with various poses, but they rely on high-level models that might fail in complex scenarios, restricting the success of the end results. Besides, high-level inference is time consuming and, thus, these methods usually are computationally expensive. The object segmentations, where each pixel is labelled with the identifier of a particular object, were used to create class segmentations, where each pixel is assigned a class label. These were provided to encourage participation from class-based methods, which output a class label per pixel but which do not output an object identifier, e.g. do not segment adjacent objects of the same class. Participants' results were submitted in the form of class segmentations, where the aim is to predict the correct class label for every pixel not labelled in the ground truth as "void".

PROPOSED METHOD

2.1 In Thresholding: Pixels are allocated to categories according to the range of values in which a pixel lies. shows boundaries which were obtained by thresholding the muscle fibres image. Pixels with values less than 128 have been placed in one category, and the rest have been placed in the other category. The boundaries between adjacent pixels in different categories have been superimposed in white on the original image. It can be seen that the threshold has successfully segmented the image into the two predominant fiber types.

2.2 In edge-based segmentation: An edge filter is applied to the image, pixels are classified as edge or non-edge depending on the filter output, and pixels which are not separated by an edge are allocated to the same category. Fig shows the boundaries of connected

regions after applying Prewitt's filter and eliminating all non-border segments containing fewer than 500 pixels. (More details will be given).

2.3 region-based segmentation: Algorithms operate iteratively by grouping together pixels which are neighbors and have similar values and splitting groups of pixels which are dissimilar in value. The boundaries produced by one such algorithm, based on the concept of watersheds.

2.3 Skin Color Segmentation: Among various low facial features such as edge, shape, skin color and texture; skin color is prominent tool for extracting face region due to its fast processing and ease of implementation. Although color processing is advantageous but sensitive to following conditions which are discussed by Ukil Yan et al.and Nidhi Tiwari et al.:

2.3.1 Illumination conditions: A change in the spectral distribution and the illumination level of light source (indoor, outdoor, highlights, shadows, color temperature of lights)

2.3.2 Camera characteristics: The color reproduced by a CCD camera is dependent on not only the illumination conditions but also the spectral sensitivities of a camera sensor.

2.3.3 Ethnicity: Skin color varies according to ethnic groups.

2.3.4 Individual characteristics: Individual characteristics such as age, sex and body parts affect the skin color.

Flow of Proposed Work

The first step of dissertation is to take RGB image as input to system. This image is pre-processed by converting from RGB to appropriate color models. After this conversion, we have segmented image in two parts as skin region and non skin region by applying thresholds for each channel of model. The threshold values come from experimentation of histograms. Thus skin region is segmented. For smooth skin area, morphological operations such as erosion and dilation are used.

The 4-point and 8-point connectivity is checked on white pixels to segment face region from image. To bound face in image with rectangle, height to width ratio is applied. This ratio avoids false detections. At last, image of face with bounding box is displayed.

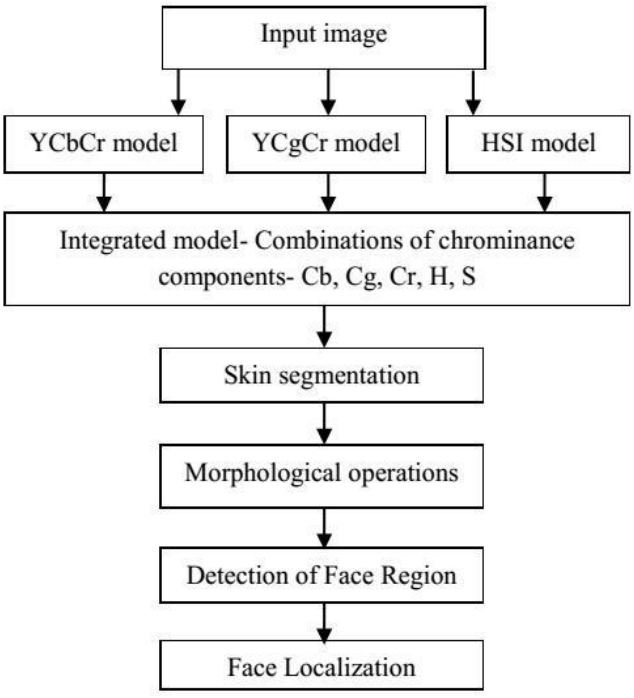

Fig. Flow of proposed work

Integration of Color Models Different combinations of chrominance components of most popular color models and their threshold values which are used for skin segmentation, shown in Table -1 These threshold values are calculated from histogram processing.

3.1 Categories of Segmentation Techniques: Two technique used namely Edge-Based and Region Based Segmentation. Both can be based on following

3.1.2 Discontinuity: It means to partition an image based on immediate changes in intensity, this includes image segmentation algorithms like edge detection.

3.1.2 Similarity: It means to partition an image into regions that are similar according to a set of predefined criterion. This includes image segmentation algorithms like thresholding, region growing, region splitting and merging.

3.2 Edge-Based Segmentation: An edge is a set of connected pixels that is lying on the boundary between two regions that differ in grey value. The pixels on the edge are called edge point. Edge-Based segmentation is also called as a Boundary based methods. It is used for finding discontinuities in gray level images. It is the best approach for detecting meaningful discontinuities in the gray level. Two types of Edge-Based Segmentation may use namely following.

3.2.1 Parallel Edge Detection: In parallel edge detection technique decide of whether or not a set of points are on an edge is independent. There are different types of parallel differential operators such as first difference operators and the second difference operator. The key difference between these operators is the weights allocated to each element of the mask.

3.2.2 Sequential Edge Detection: In Sequential edge detection technique, the result at a point is dependent on the result of the before examined points. The act of a sequential edge detection algorithm will depend on the choice of a good initial point, and it is not easy to define termination criteria.

3.3 Region-based Segmentation: Region based segmentation techniques split the entire image into sub regions depending on some rules. Rules like all the pixels must have the same gray level. Region-based segmentation methods attempt to group regions allowing to common image properties. Edge based methods partition an image based on rapid changes in intensity nearby edges whereas region based methods, partition an image into regions that are related according to a set of predefined criteria.

3.3.1 Region Growing: Region growing is a procedure that group's pixels in whole image into sub regions based on predefined standard. Region Growing is used to group a collection of pixels with related properties form a region.

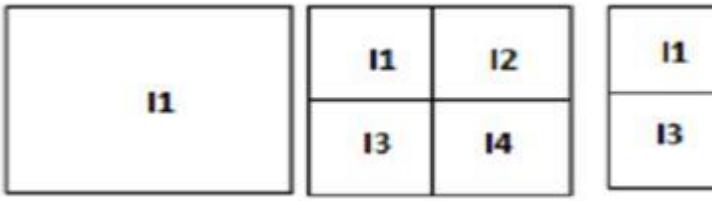

Fig 3.3 Example of images for a region growing

BLOCK DIAGRAM

We overview the method here for the upper-body case, where there are 6 parts: head, torso, and upper/lower right/left arms. The method is also applicable to full bodies, as demonstrated.

A recent and successful approach to 2D human tracking in video has been to detect in every frame, so that tracking reduces to associating the detections. We adopt this approach where detection in each frame proceeds in three stages, followed by a final stage of transfer and integration of models across frames.

In our case, the task of pose detection is to estimate the parameters of a 2D articulated body model. These parameters are the (x, y) location of each body part, its orientation θ, and its scale. Assuming a single scale factor for the whole person, shared by all body parts, the search space has $6 \times 3 + 1 = 19$ dimensions. Even after taking into account kinematic constraints (e.g. the head must be connected to the torso), there are still a huge number of possible configurations.

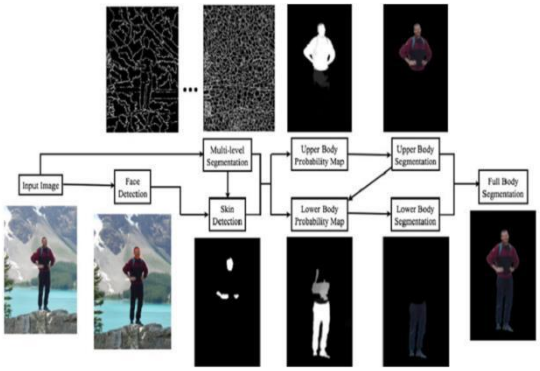

Fig.1 Overview of the methodology

Face detection guides estimation of anthropometric constraints and appearance of skin, while image segmentation provides the image's structural blocks. The regions with the best probability of belonging to the upper body are selected and the ones that belong to the lower body follow.

Approach overview: Since at the beginning we know nothing about the person's pose, clothing appearance, location and scale in the image, directly searching the whole space is a time consuming and very fragile operation (there are too many image patches that could be an arm or a torso!). Therefore, in our approach the first two stages use a weak model of a person obtained through an upper-body detector generic over pose and appearance. This weak model only determines the approximate location and scale of the person, and roughly where the torso and head should lie. However, it knows nothing about the arms, and therefore very little about pose. The purpose of the weak model is to progressively reduce the search space for body parts. The next two stages then switch to a stronger model, i.e. a pictorial structure describing the spatial configuration of all body parts and their appearance. In the reduced search space, this stronger model has much better chances of inferring detailed body part positions.

Skin Detection: Among the most prominent obstacles to detecting skin regions in images and video are the skin tone variations due to illumination and ethnicity, skin-like regions and the fact that limbs often do not contain enough contextual information to discriminate them easily. In this study, we propose combining the global detection technique with an appearance model created for each face, to better adapt to the corresponding human's skin color (Fig. 6.2). The appearance model provides strong discrimination between skin and skin-like pixels, and segmentation cues are used to create regions of uncertainty. Regions of certainty and uncertainty comprise a map that guides the Grab Cut algorithm, which in turn outputs the final skin regions. False positives are eliminated using anthropometric constraints and body connectivity

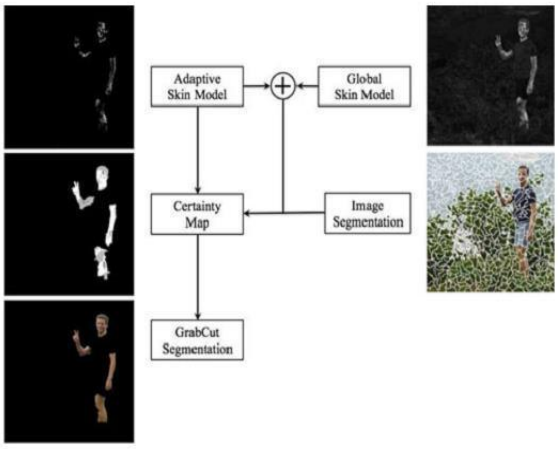

Fig. Skin detection examples

The adaptive model in general focuses on achieving a high score of true positive cases. However, most of the time it is too "strict" and suppresses the values of many skin and skin-like pixels that deviate from the true values according to the derived probability distribution. At this point, we find that an influence of the skin global detection algorithm is beneficial because it aids in recovering the uncertain areas. Another reason we choose to extend the skin detection process is that relying solely on an appropriate color space to detect skin pixels is often not sufficient for real-world applications.

Fig. Skin detection algorithm

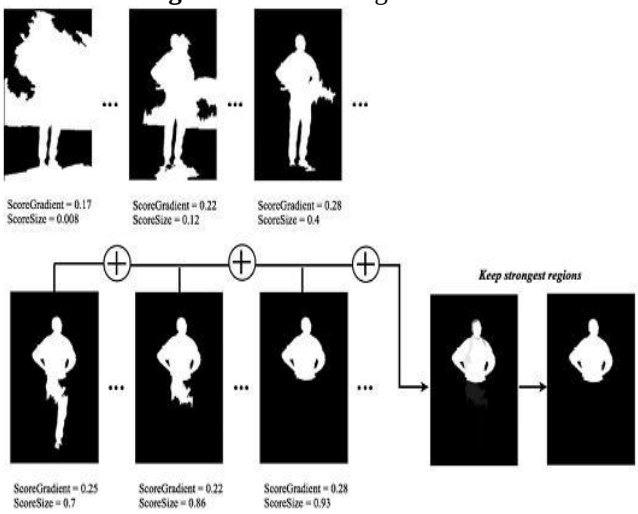

Fig. Thresholding of the aggregated potential torso images and final upper body mask. Note that the masks in the top row are discarded.

The obvious step is to threshold the aggregated potential torso images in order to retrieve the upper body mask. In most cases, hands or arms' skin is not sampled enough during the torso searching process, especially in the cases, where arms are outstretched. Thus, we use the skin masks estimated during the skin detection process, which are more accurate than in the case they were retrieved during this process, since they were calculated using the face's skin color, in a color space more appropriate for skin and segments created at a finer level of segmentation. These segments are superimposed on the aggregated potential torso images and receive the highest potential.

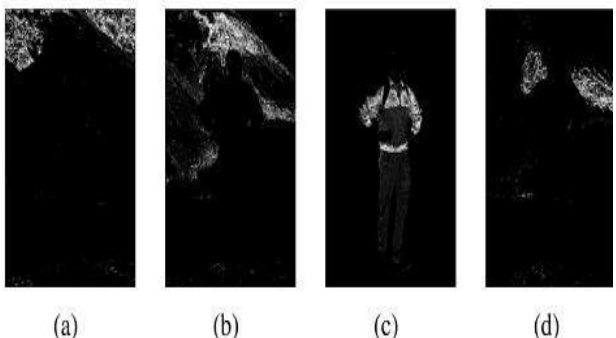

Fig. Example of similarity images for random segments

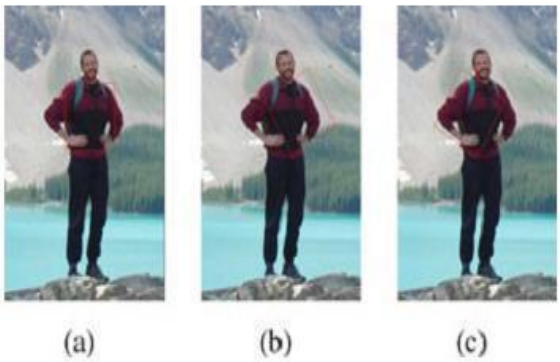

Fig. Masks used for torso localization

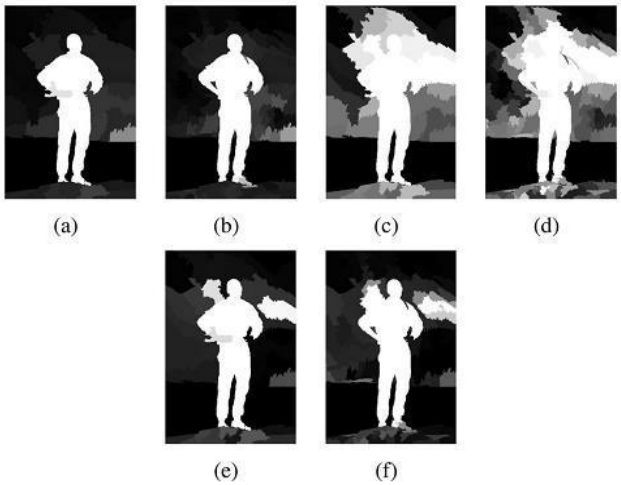

Fig.6.6. Segments with potential of belonging to torso. (a), (b) For segmentation level 1 and 2 and torso mask at 0 ◦. (c), (d) For segmentation level 1 and 2 and torso mask at 30 ◦. (e) (f) For segmentation level 1 and 2 and torso mask at −30 ◦.

Fig. Example legs mask for φright = 0 and φleft = 0

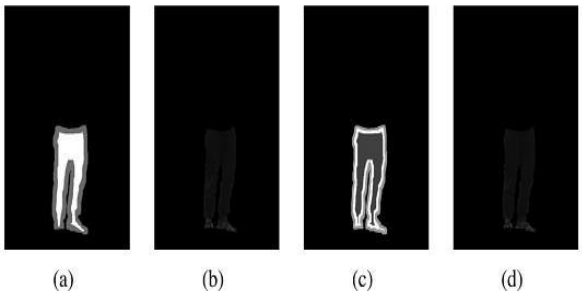

Fig. Example of foreground/background certainty maps and segmentations for (a) and (b) Grab Cut and (c) and (d) Grow Cut.

SOFTWARE SPECIFICATIONS AND FRAMEWORK

Software Specifications
Operating System : Linux
MAT LAB, C, FORTRAN

RESULTS

Instead of using a simple or even adaptive thresholding, we use a multiple level thresholding to recover the regions with strong potential according to the method described, but at the same time comply with the following criteria:

1) they form a region size close to the expected torso size (actually bigger in order to allow for the case, where arms are outstretched), and
2) the outer perimeter of this region overlaps with sufficiently high gradients. The distance of the selected region at threshold t (Region) to the expected upper body size (Exp Upper Body Size) is calculated as follows:

ScoreSize =

where Exp Upper Body Size = 11 × PL2 . The score for the second criterion is calculated by averaging the gradient image (Grad Im) responses for the pixels that belong to the perimeter (Region) of Region as results. Specifically, we set a final threshold, which allows only regions that have survived more than 20% of the accumulation process in the final mask for the UBR. This process is performed for every initial torso hypothesis; therefore, in the end, there are three corresponding aggregate masks, out of which the one that overlaps the most with the initial torso mask and obtains the highest aggregation score is selected. The aggregation score shows how many times each pixel has appeared in the accumulation process, implicitly implying its potential of belonging to the true upper body segment.

CONCLUSION

We presented a novel methodology for extracting human bodies from single images. It is a bottom-up approach that combines information from multiple levels of segmentation in order to discover salient regions with high potential of belonging to the human body. The main component of the system is the face detection step, where we estimate the rough location of the body, construct a rough anthropometric model, and model the skin's color. Soft anthropometric constraints guide an efficient search for the most visible body parts, namely the upper and lower body, avoiding the need for strong prior knowledge, such as the pose of the body.

ScoreGrad =

REFERENCES

[1] Solomon, C.J.; Breckon, T.P. Thresholding starts with zero and becomes increasingly stricter at small steps (0.02). In each thresholding level, the largest connected component is rated, and the masks with Score Grad > 0.05 and Score Size > 0.6 are accumulated to a refined potential image (see in Fig. 6.8). Incorporation of this a priori knowledge to the thresholding process aids the accentuation of the true upper body regions (UBR). Accumulation of surviving masks starts when Score Size > 0.6 and resulting masks after this point will keep getting closer monotonically to the expected region size. Accumulation ends when Score Size drops below 0.6. The rationale behind this process is to both restrict and define the thresholding range and focus the interest to segments with high potential of forming the upper body segment. The aggregate mask (Aggregate Mask) can now be processed easily and produce more meaningful (2010). Fundamentals of Digital Image Processing: A Practical Approach with Examples in Matlab. Wiley-Blackwell. doi: 10.1002/9780470689776. ISBN 0470844736.

[2] Rafael C. Gonzalez; Richard E. Woods; Steven L. Eddins (2004). Digital Image Processing using MATLAB. Pearson Education. ISBN 978-81-7758-898-9.

[3] V. Ferrari, M. Marin-Jimenez, and A. Zisserman, "Progressive search space reduction for human pose estimation," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2008, pp. 1–8.

[4] M. P. Kumar, A. Zisserman, and P. H. Torr, "Efficient discriminative learning of parts-based models," in Proc. IEEE 12th Int. Conf. Comput. Vis., 2009, pp. 552–559.

[5] V. Delaitre, I. Laptev, and J. Sivic, "Recognizing human actions in still images: A study of bag-of-features and part-based representations," in Proc. IEEE Brit. Mach. Vis. Conf., 2010.

[6] A. Gupta, A. Kembhavi, and L. S. Davis, "Observing human-object interactions: Using spatial and functional compatibility for recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 10, pp. 1775–1789, Oct. 2009.

[7] B. Yao and L. Fei-Fei, "Grouplet: A structured image representation for recognizing human and object interactions," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2010, pp. 9–16.

[8] P. Buehler, M. Everingham, D. P. Huttenlocher, and A. Zisserman, "Long term arm and hand tracking for continuous sign language TV broadcasts," in Proc. 19th Brit. Mach. Vis. Conf., 2008, pp. 1105–1114.

[9] A. Farhadi and D. Forsyth, "Aligning ASL for statistical translation using a discriminative word model," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog., 2006, pp. 1471– 1476.

[10] L. Zhao and L. S. Davis, "Iterative figure-ground discrimination," in Proc. 17th Int. Conf. Pattern Recog., 2004, pp. 67–70.

[11] L. Grady, "Random walks for image segmentation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, no. 11, pp. 1768–1783, Nov. 2006.

[12] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut: Interactive foreground extraction using iterated graph cuts," ACM Trans. Graph., vol. 23, no. 3, pp. 309–314, Aug. 2004.

[13] V. Gulshan, C. Rother, A. Criminisi, A. Blake, and A. Zisserman, "Geodesic star convexity for interactive image segmentation," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2010, pp. 3129–3136.
[14] Y. Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in ND images," in Proc. IEEE 8th Int. Conf. Comput. Vis., 2001, pp. 105–112.
[15] M. P. Kumar, P. H. S. Ton, and A. Zisserman, "Obj cut," in Proc. IEEE Comput. Soci. Conf. Comput. Vision Pattern Recog., 2005, pp. 18–25.
[16] S. Li, H. Lu, and L. Zhang, "Arbitrary body segmentation in static images," Pattern Recog., vol. 45, no. 9, pp. 3402–3413, 2012.
[17] L. Huang, S. Tang, Y. Zhang, S. Lian, and S. Lin, "Robust human body segmentation based on part appearance and spatial constraint," Neurocomputing, vol. 118, pp. 191–202, 2013.
[18] P. F. Felzenszwalb and D. P. Huttenlocher, "Pictorial structures for object recognition," Int. J. Comput. Vis., vol. 61, no. 1, pp. 55–79, 2005.
[19] D. Ramanan, "Learning to parse images of articulated bodies," Adv. Neur. Inf. Process. Sys., pp. 1129–1136, 2006.
[20] M. Eichner and V. Ferrari, "Better appearance models for pictorial structures," in Proc. Brit. Mach. Vis. Conf., 2009.
[21] Y. Bo and C. C. Fowlkes, "Shape-based pedestrian parsing," in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recog., 2011, pp. 2265– 2272.
[22] Z. Hu, G. Wang, X. Lin, and H. Yan, "Recovery of upper body poses in static images based on joints detection," Pattern Recog. Lett., vol. 30, no. 5, pp. 503–512, 2009.
[23] J. Malik, S. Belongie, T. Leung, and J. Shi, "Contour and texture analysis for image segmentation," Int. J. Comput. Vis., vol. 43, no. 1, pp. 7–27, 2001.
[24] M. Yao and H. Lu, "Human body segmentation in a static image with multiscale superpixels," in Proc. 3rd Int. Conf. Awareness Sci. Technol., 2011, pp. 32–35.
[25] Y. Hu, "Human body region extraction from photos," in Proc. Mach. Vis. Appl., 2007, pp. 473–476.